C1

SOUTHWEST FISHERIES SCIENCE CENTER

NATIONAL MARINE FISHERIES SERVICE - HONOLULU LABORATORY - 2570 DOLE STREET HONOLULU, HI 96822-2396

APRIL 2001

LIBRARY

march 28, 2001

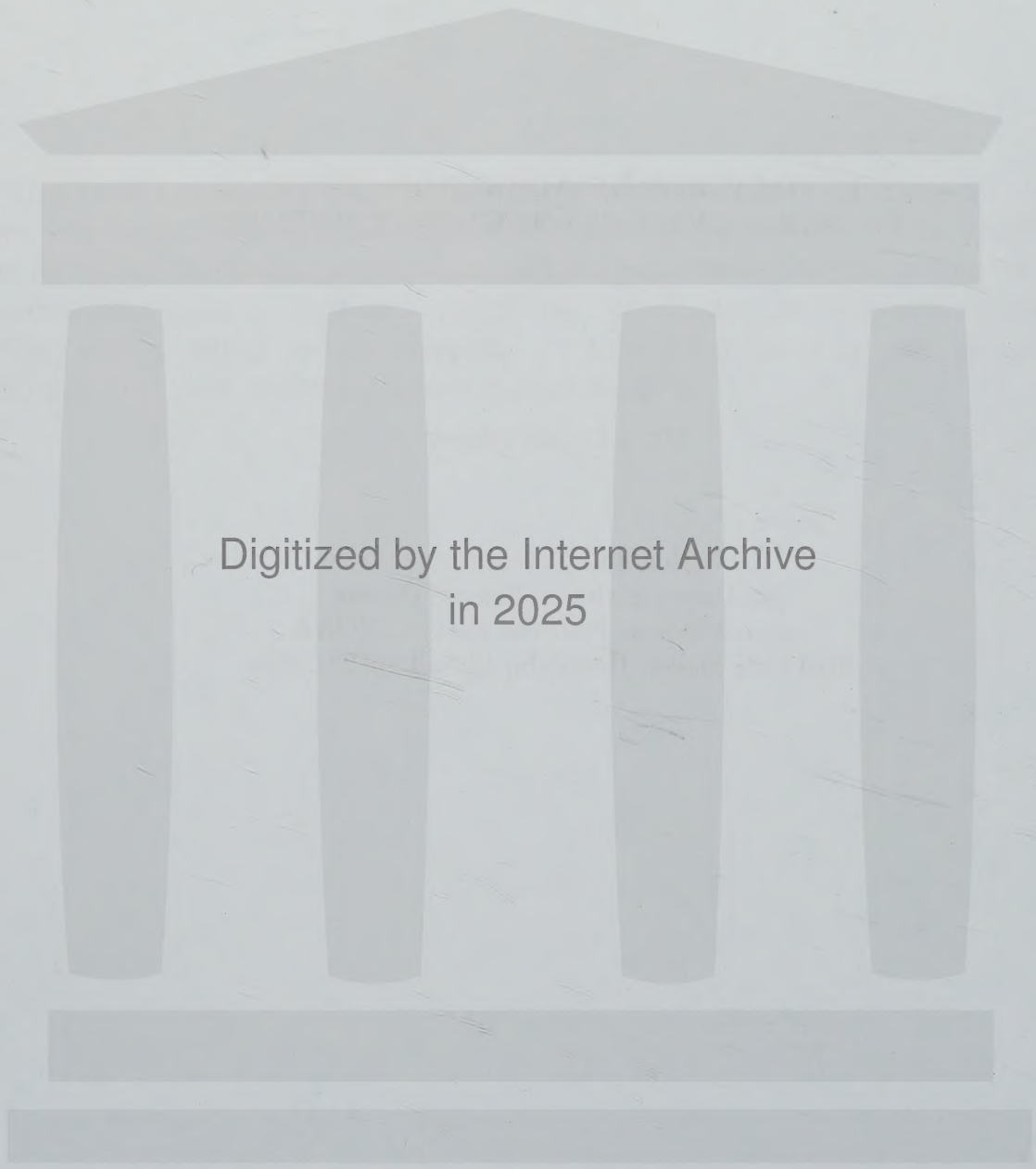
Marine Fisheries Service
2570 Dole Street
Honolulu, Hawaii

ESTIMATION OF ALBATROSS TAKE IN THE HAWAIIAN LONGLINE FISHERIES

MARTI L. McCRACKEN

Honolulu Laboratory
Southwest Fisheries Science Center
National Marine Fisheries Service, NOAA
Honolulu, Hawaii 96822-2396

Administrative Report H-01-03



Digitized by the Internet Archive
in 2025

https://archive.org/details/administrative-report-h_2001-04_3

1. INTRODUCTION

This report explains the statistical methodology used to estimate total take of black-footed albatross (*Phoebastria nigripes*) and Laysan albatross (*Phoebastria immutabilis*) for all trips landing in Hawaii by the Hawaiian-based longline fishery entered into the logbook database. Although takes of seabirds and sea turtles are recorded in the logbooks, this information is viewed as unreliable. Based on biological opinions issued under the Endangered Species Act, an observer program was established in 1994 to collect data for monitoring sea turtle takes; data to monitoring seabird takes are also collected. The observer program places trained observers on randomly selected trips of the longline fishery. Since a list of trips does not exist until the end of the year, it is assumed that all, or nearly all, Hawaii-based longline fishing trips landing in Hawaii have a probability greater than zero of being sampled. If a trip has zero probability of being sampled, it is assumed that the reason for being excluded is not related to seabird take and excluding it does not introduce bias. We also assume that the presence of an observer does not change the behavior of the fishermen. During observed trips, the number of seabird takes by species, the condition of the seabirds, other species of concern, and possible explanatory variables are recorded by the observer for each longline set. A seabird take is defined as an interaction between a seabird and anything related to the activity of fishing, and it usually implies that the seabird became entangled in the line or was caught on a hook. In this report, the observer and logbook records from 1994 through 1999 were used to estimate total black-footed and Laysan albatross takes by year. Since sampling did not begin until late February 1994, we extrapolated from our sampled population to predict take for the first couple of months of 1994.

Model-based predictors, instead of sample-based estimators, have been used to estimate total albatross takes by species since 1996. Whereas sample-based estimators assume sampling probabilities to raise observed total takes to the fleet level, model-based predictors assume a statistical model of albatross takes. Sample-based estimators have the advantage of being basically free of assumptions concerning the structure of the target population and the characteristics being estimated, but they are typically less efficient than model-based predictors. Model-based predictors are dependent on the model, and departures from the assumed model are likely to introduce bias. We have chosen to continue with model-based prediction because of insufficient information on sampling probabilities, and with only 5% of trips being sampled a year, we would expect sample-based estimators to be imprecise.

We present the methods used for modeling and prediction and discuss the results. Section 2 describes the data structure, and the methodology used for developing the prediction model is explained in Section 3. Section 4 describes how total takes were estimated and prediction intervals approximated. The final prediction models and estimated total takes are presented in Section 5.

2. Data structure

Data from the observer program are hierarchical, with trip as our independent observational unit and sets within a trip defined as subunits. Because of this structure, two types of stochastic dependence among sets from the same trip may exist: (1) takes from

sets within a trip may be more closely related than takes across trips and (2) takes from sets close together in time and space within a trip may be more closely related. If take is modeled at the trip level, we could assume independent observations, but information is lost in the explanatory variables. For example, latitude and longitude are recorded for each set and these variables would need to be summarized if modeling at the trip level. Hence, modeling at the set level is preferred, but results can be misleading if the hierarchical structure of the data is ignored. Therefore, if explanatory variables do not explain the dependence among sets, the dependence should be modeled. Trips by the same vessel may also be dependent, but without sufficient replication of vessels we lack the information to model this dependence.

Examination of the observer data suggests that black-footed and Laysan albatross take from sets within the same trip are probably not independent. For black-footed albatross, 184 trips had 0 observed takes, but out of the 82 trips with positive takes, 39 had positive takes in more than one set. The three highest black-footed albatross takes recorded for a trip were 50, 33, and 21; these takes were accumulated over several sets. Takes for the other trips were less than 18. For Laysan albatross, 177 trips had 0 observed takes, but out of the 57 trips with positive takes, 22 had positive takes in more than one set. The three highest Laysan albatross takes recorded for a trip were 85, 32, and 29; these takes were accumulated over several sets. Takes for the other trips were less than 20. If this dependence structure could be successfully modeled parametrically, we would expect a more efficient predictive model, but further work is needed to determine if there is an 'appropriate' model. Therefore, albatross takes were modeled at the set level, but the likely dependence among observations was considered when selecting and fitting the models and estimating uncertainity.

Prediction models for albatross takes were constructed using observer data and corresponding logbook data. The response variable was take recorded by the observer at set level. This type of data is frequently referred to as count data; i.e., we counted the number of times an event, albatross take, occurred. Albatross take is a discrete variable, typically the number of positive takes was 1, 2, 3, or 4, but up to 12 takes were observed during a set. Figures 1 and 2 show approximate geographical positions of all observed sets and those sets with positive takes, represented by stars. For both albatross species, the set-level data contained a high percentage of zero takes: 91.1% for black-footed, 94.3% for Laysan. If the average count is sufficiently large enough that counts, or a transformation of them, are approximately normal, robust modeling methods based on the Gaussian distribution can be used. Our average take per set was small even in the northern latitudes where most takes occurred; therefore, take was modeled as a discrete random variable.

3. Developing models for albatross take

Generalized additive models (GAM) and generalized linear models (GLM) were used as exploratory tools in developing the final predictive models. When developing a prediction model, only explanatory variables that were well represented in logbooks were considered. A variable that is likely related to take is the number of albatross within the vicinity of the fishing vessel during the set. Although this variable is unknown, we did have the

estimated number of breeding pairs at relevant breeding colonies for 1998 and the set's location relative to the colonies. As a proxy for the number of albatross within vicinity of the vessel, we used the summary function

$$\sum_{s=1}^{S} n_s \exp\left(\frac{-x_s^2}{2 \cdot 10i^2}\right), \tag{3.1}$$

where s, $s = 1, \ldots, S$, represents the breeding colony for S relevant colonies, n_s is the number of estimated breeding pairs in 1998 (supplied by U.S. Fish and Wildlife Service) for the species of interest, and x_s is the distance from the set to the breeding colony. This is the same function defined and used by Kleiber (personal communications) in previous years. We tried several values of i, but finally settled on $i = 50$ for black-footed albatross and $i = 60$ for Laysan albatross as these values appeared to provide the best predictors. Table 1 lists the variables that were considered as predictors.

Basically, the selected model should be the simplest possible model that will accurately predict total take. To gain the best predictive power and safeguard against bias, the selected model should include as many predictors as possible so as not to miss anything. The problem with including as many predictors as possible is the *curse of dimensionality*, where dimensionality refers to the number of predictors. There are two problems with increasing the dimensions of a model. First, as the number of predictors increases the prediction variance increases. Second, progressively larger sample sizes are required in higher dimensions to achieve comparable accuracy. Hence, model selection requires balancing between variance and bias.

Many of the explanatory variables considered were related to each other, and if one of these variables were included in the model, including the other variables was redundant. To understand the relationships between explanatory variables, the variables were considered individually, in groups of related variables, and in sets that appeared to be good predictors. If a group of variables were associated with take, but the variables were related, careful consideration was taken not to be redundant and to select the best predictors. If a set of these variables was not clearly superior, the set of predictors selected was based on (1) minimizing measurement error, (2) assessing the distribution of each variable around its mean since prediction error will increase as we move away from the average predictor values, and (3) determining the degree to which the range of the logbook variable was covered in the observed sets since extrapolation is often less reliable than interpolation.

Because the units of measurement of different quantitative variables varied considerably, all quantitative variables were scaled according to the equation

$$x_i' = \frac{2x_i - max(\mathbf{x}) - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})}, \tag{3.2}$$

where x_i represents the i^{th} observation of variable x, and $\mathbf{x}$ represents the vector of observed values for x. This transformation should also improve the fit of the GAMs since the reliability of the GAM fit is always reduced at the endpoints of the variable's range (Hastie and Tibshirani, 1990), and the transformation will bring in the tails of the distribution.

3.1. **Generalized Linear Models.** The Poisson distribution is a natural distribution to assume for counts of rare events. For Poisson counts, the generalized linear model known as the log-linear model is applicable. The log-linear Poisson model assumes that all counts $(Y_i, i = 1, \ldots, n)$ are independent Poisson variates with mean μ_i and observed value $y_i = \mu_i + \epsilon_i$, where ϵ_i is the residual or 'error' and n is the number of observations. The mean is modeled as

$$log(\mu_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij},$$

where $(x_{i1}, \ldots, x_{ip})$ are the explanatory variables and the βs are (usually) unknown parameters to be estimated. The right side of the equation is referred to as the linear predictor. The function that connects the linear predictor to the mean μ of Y, in this case the natural logarithmic function (log), is known as the link function. The link function that transforms the mean to the canonical parameter of the exponential family is called the canonical link; for the Poisson distribution, the canonical link is the log function. An attractive property of the canonical links is that $X^T Y$ (in vector notation where X is the matrix of predictors) is a sufficient statistic equal in dimension to β (McCullagh and Nelder, 1989). Although the canonical link leads to desirable statistical properties of the estimated model, particularly for small samples, there is generally no a priori reason why systematic effects in a model should be additive on the scale given by that link. Typically maximum likelihood estimates (MLE) are used as parameter estimates, and these are typically derived using iterative weighted least squares or the Newton-Raphson method.

Unlike the normal distribution that is completely determined by its two parameters, the mean μ and the variance σ^2, the relationship between the mean and the variance for a Poisson variate is a fixed relationship. Given a Poisson distribution with parameter λ, this relationship is $\mu = \sigma^2 = \lambda$. Hence, determining the mean of the distribution implicitly determines its variance, and there is no flexibility. This is in contrast to linear regression models where the response is assumed to be normally distributed and where the variance parameter and mean can be estimated separately; thus, any constant degree of variability about the mean is accommodated in the fitted model. For these reasons, a Poisson distribution may seem suitable for a response, but on close inspection, the variance does not appear to equal the mean as assumed. If the variance is larger, the data exhibit overdispersion; if it is smaller, the data exhibit underdispersion. Overdispersion is more common and typically arises as a result of some sort of clustering or clumping process. With bird counts, the data are expected to exhibit overdispersion if the birds tend to aggregate and underdispersion if the population is spread very evenly over its range. Overdispersion can also arise if the occurrence of an event increases or decreases the subsequent probability of another event. In terms of albatross takes, overdispersion can arise if albatross are attracted to a vessel that they have learned is a food source, and their presence attracts other birds; as the number increases, the probability of one or more takes likely increases. Another way in which overdispersion may arise is when there is inter-trip variability; i.e., the number of takes for sets within a trip might be Poisson with mean λ_t, but λ_t varies between trips. If there is a relationship between take and the unmeasured

ability of the fishing crew, then sets within a trip may be related by a common Poisson mean, but this mean will vary between trips. Failure to acknowledge overdispersion can lead to serious underestimation of standard errors and hence to misleading inferences about the form of the linear predictor. It is important to consider overdispersion when modeling albatross take at the set level since our independent unit is the trip.

Various methods could be adopted for dealing with the problem, but one straightforward approach is quasi-likelihood estimation. Quasi-likelihood assumes a functional relationship between the mean and the variance; that is, it assumes that the variance has the form $Var(Y) = V(\mu)$ for some chosen function V. One commonly assumed form for Poisson data is $Var(Y) = \phi\mu$, where $\phi > 1$ represents overdispersion and $\phi < 1$ underdispersion. Even relatively substantial errors in the assumed functional form of $Var(Y)$ generally have only a small effect on the conclusions (McCullagh and Nelder, 1989). An advantage of quasi-likelihood estimation is that point estimates do not depend on the value of σ^2. Using quasi-likelihood methods, ϕ cancels out in our estimating equations, so that regression parameters remain the same as if ϕ were equal to 1. The parameter ϕ is typically called the dispersion parameter, and an estimate of ϕ can be used in estimating the standard errors for the $\hat{\beta}$s. However, quasi-likelihood estimators are not MLE without the additional assumption that the responses are Poisson distributed. Nevertheless, McCullagh (1983) showed they have similar properties as MLE. Under quite general conditions, they are consistent and asymptotically normal. When quasi-likelihood estimators are not MLE, Cox (1983) and Firth (1987) showed that they still retain relatively high efficiency as long as the degree of overdispersion is moderate. However, for the estimation of ϕ, quasi-likelihood does not behave like a log-likelihood.

For Poisson data, a conventional estimate of ϕ is

$$\hat{\phi} = \frac{1}{n-p}\sum_i \frac{(y_i - \mu_i)^2}{\mu_i} = \frac{\chi^2}{(n-p)},$$

where χ^2 is the generalized Pearson statistic. However, for the albatross data this estimator was unsatisfactory because of the number of μ_i very near 0. For example, there is only 1 observed black-footed albatross take south of $20°N$; the other 951 observed sets below this latitude had zero takes. Estimation of the dispersion for means very close to zero becomes difficult because of the sample size required. If we ignore this problem and include these sets when estimating overdispersion, the estimated dispersion parameter is 1.07, and we would likely conclude there is no evidence of overdispersion. However, if these values are excluded, the estimated dispersion parameter is 1.49, and this value is likely still underestimating overdispersion. Therefore, this estimate was not used for the dispersion parameter.

Alternately, the counts can be modeled as negative binomial variates. If we consider the number of takes for sets within the same trip as Poisson variates with mean λ_t, where λ_t is a gamma variate with mean μ and variance $\mu^2\nu$, then takes have a negative binomial distribution with mean μ and variance $\mu + \mu^2/\nu$. The term μ can be thought of as the Poisson variance component and μ^2/ν as the extra component arising from mixing the Poisson distribution with a gamma distribution to obtain the negative binomial (McCullagh and Nelder, 1989). S-PLUS fits the negative binomial model by an alternating

iteration process (Venables and Ripley, 1999). For a given ν, the GLM is fitted using the same process as for fitting a Poisson model. For fixed means, ν is estimated using score and information iterations. The initial value of ν is the moment estimator after an initial fitting of a Poisson model. The two are alternated until convergence of both. The canonical link for the negative binomial is $log(\mu/(\mu+\nu))$, and if ν is unknown, the use of the canonical link is problematical. The links currently available for the negative binomial in S-PLUS are the log, square root, and identity link.

3.2. **Generalized Additive Models.** A limitation of GLMs is that they restrict the parameters relating the predictors to the link function to be of linear form. A more flexible model would be one that does not make any assumptions about the shape of the functions relating the predictors to the link function. This is the motivation behind the development of GAMs. Basically GAMs replace the GLM linear predictor, $\eta = \sum_{j=1}^{p} \beta_j x_j$, with a flexible additive function $\eta = \sum_{j=1}^{p} s_j(x_j, d_j)$, where $s_j(x_j, d_j)$ is an unspecified smooth function of x_j, and d_j is the degrees of freedom of the smoother. GAMs still assume additivity in the predictors, but they allow the smooth function to take on any shape ranging from a straight line to nonparametric curves of increasing complexity. Similar to GLMs, the model specification is completed by a variance function, $Var(Y) = V(\mu)$. Categorical variables are permitted in the GAM environment and are expressed in the same manner as they are in GLMs. An advantage of working in the GAM environment is that linear, smooth, and step functions of the predictors can be compared.

GAMs are fitted by estimating the smoothing functions from the data. An iterative algorithm that uses backfitting in combination with local scoring is probably the most widely used method for fitting a GAM. This is the algorithm used by S-PLUS. The most critical part of this algorithm is the backfitting that solves the system of normal equations. One disadvantage of the backfitting procedure is that the bias and variance of the smoothers cannot be derived theoretically except in special cases. Currently it is believed that resampling techniques are the better choice for estimating standard errors and approximating confidence intervals. Additionally, the convergence behavior of the backfitting and similar iterative procedures is not well understood. Recently methods, such as the marginal-integration method, that avoid iterative techniques in fitting additive models have been developed that are theoretically interesting but too new to understand their value in analyzing real data. Backfitting works better at boundary points and under data sparseness than the marginal method, but both of these algorithms run into severe problems with increasing collinearity or concurvity (concurvity refers to nonlinear dependence) among explanatory variables. The relaxed iterative projection algorithm (Schimek and Turlach, 2000) is a newer algorithm that has the potential to perform better under these circumstances, but it also is too new to understand its statistical behavior. Another concern when modeling take at the haul level is that most methods for fitting GAMs assume independent errors, and estimates of s can be adversely affected if the errors are dependent. A major drawback with GAMs is the inability to include, estimate, and test for interactions between nonlinear smooth predictors, although there have been some very recent advancements in this area. The methodology for fitting GAMs to negative binomial variates has just recently been developed (Thurston et al., 2000),and

software is not generally available. See Schimek and Turlach (2000) for more details and references concerning fitting GAMs.

With almost all estimation methods for GAMs, the degree of smoothing for each dimension must be specified. The minimal degree of smoothing assumes a constant slope and is just a straight line $s(x) = \beta x$. The maximal is when $s(x)$ fits the data perfectly; the smoother goes through all points by connecting successive pairs of points with lines having their own individual slopes. At these two extremes, the right side of the equation can be expressed as a linear predictor and estimated using a GLM. Between these extremes, the function s is usually not specified but is estimated nonparametrically from the data by means of a scatterplot smoother. The shape of the function is therefore determined by the data rather than being restricted to a parametric form. The amount of smoothing in a GAM is often determined by the prespecification of the (equivalent) degrees of freedom. As the degrees of freedom is increased, the function s gains flexibility and becomes 'rougher,' displaying more hills and valleys with more complex shapes. As with any model building, we want the simplest possible model that will achieve our required purpose. In the context of smoothing, we want the fewest degrees of freedom required. Because of the lack of theory and adequate algorithms, how to best select a smoothing parameter by a data driven method remains a troubling issue (Schimek and Turlach, 2000). In the univariate situation (one explanatory variable), generalized cross-validation, the improved Akaike information criterion (Hurvich et al., 1998), or comparable techniques could be used. In principle, these methods could be extended to the multivariate situation, but for the data-driven methods finding a global minimum is difficult.

There are several different types of scatterplot smoothers. However, the methodology used to fit the smoother is generally more critical than the type of smoother specified. Methods for fitting GAMs are often only applicable to certain types of smoothers. For example, the marginal-integration method is restricted to local polynomial and kernel fits. The cubic smoothing spline fitted by satisfying a penalized log-likelihood criterion using the backfitting algorithm in combination with local scoring is probably the most widely available methodology used to fit GAMs. The intention of the penalized log-likelihood criterion is to optimize the fit while penalizing roughness to some prespecified extent.

For GLMs, model evaluation and selection are supported by asymptotic distribution theory, but this generally is not true for GAMs. For example, the deviance statistic for a GLM has an asymptotic χ^2-distribution, but for a GAM it does not. However, the deviance and related likelihood statistics remain a useful tool for evaluating and selecting models, although extreme caution should be taken when relying on GLM asymptotic theory. Hastie and Tibshirani (1990) provide empirical evidence that supports using the χ^2-distribution when evaluating the deviance. In summary, model selection remains a weak point in this class of models.

3.3. **Fitting and Selecting the Models.** All models in this analysis were fitted using the statistical software S-PLUS (Statistical Sciences, Inc., 1993). We modeled takes using GAMs with a log link assuming an overdispersed Poisson distribution and GLMs with a log link assuming a negative binomial and overdispersed Poisson distribution. The β's in

the linear predictor of the GLMs were estimated by maximum or quasi-likelihood. GAMS were fitted with cubic smoothing splines chosen to satisfy the penalized log-likelihood criterion.

To develop a subset of plausible linear predictors, stepwise selection based on the generalized information criterion was used. The generalized information criterion in this context is defined as

$$GIC = D - \alpha p\hat{\phi},$$

where D is the deviance, p is the number of parameters in the model, and α is either constant or a function of n, the number of observations. Two of the most commonly used forms of the GIC are Akaike's information criterion (AIC) where $\alpha = 2$ and Bayesian information criterion (BIC) where $\alpha = log(n)$. S-PLUS stepwise selection within the GAM framework is based on AIC. The theory behind GIC, including AIC, assumes MLE and independent observations and is a function of the dispersion parameter. However, (1) estimates of smoothers are not generally true MLEs, although many of the methodologies used to fit smoothers are related to MLE; (2) assuming that sets within trips are independent is questionable; (3) we do not have a good estimate of the dispersion parameter; and (4) AIC may not be the optimal form of the GIC for our data. The GIC with $\alpha = 2$ (AIC) leads to a ranking of the model in order of the estimated mean square error of prediction, and under many situations, 2 is the optimum value for α. But as n increases there will eventually be a point where larger values of α are optimum (Atkinson, 1980). Also, larger values of α are indicated if the prediction problem is ill conditioned; the matrix X^TX is nearly singular (Atkinson, 1980). Along these same lines, as $\mu \to 0$ there is likely a point where larger values for α are optimum since $\mu = 0$ is not estimable. When initially fitting the models and using AIC ($\alpha = 2$) and assuming a Poisson distribution ($\phi = 1$), model diagnostics indicated that AIC tended to overfit. Because we have over 3000 observed sets and a high level of collinearity and concurvity among explanatory variables, a larger value of α is likely required for optimum model selection. Additionally, because we are fitting the models treating sets as independent observations, care must be taken not to overfit the model. BIC adjusts for the sample size, and for the analysis in this paper, BIC is the GIC with $\alpha = log(3107) \doteq 8$. However, even BIC in this circumstance may not be optimum. Atkinson (1981) suggests that the range $\alpha = 2$ to 6 may provide a set of plausible initial models for further analysis, but since $\alpha = 8$ is BIC, we extended this range. Another reason AIC may be overfitting is that there is overdispersion, and variability is being underestimated.

When the GIC is used to select the degree of smoothing, for standard data, AIC tends to undersmooth (Hurvich et al., 2000). This suggests that the optimal level of α for selecting the smoothing parameter using GIC may be less than the optimal level for selecting the predictor variables. Work in progress with simulated data supports this, but more research is required before making conclusions.

For all these reasons, stepwise selection within S-PLUS was used only as an aid in determining the final predictive model. Since we do not have a good estimate of ϕ and a higher value for α than 2 is likely optimum, the dispersion parameter was specified over a range of values from 1 to 10. The higher values for the dispersion parameter place a higher penalty on adding a predictor, so we were adjusting for α as well as overdispersion. When

using stepwise selection, we recorded the final model selected and models with similar GIC values. Under this framework, stepwise selection provided a useful tool.

To evaluate if dependence between sets was a factor in some variables being selected, we randomly withdrew one set from each trip and used stepwise selection on the remaining sets. In this instance, we used AIC since our sample size and correlation between explanatory variables were reduced. We did this 100 times and then recorded the frequency a variable was included in the selected model. This exercise proved useful in highlighting variables that were being selected because of a couple of trips that had several sets with high takes. These variables were dropped from the model as they did not appear to explain the general pattern of take levels.

Variables were introduced into the GAMs and GLMs first in similar groups and then combined. Also, since selection was dependent on the order in which predictors were introduced, the order was varied. For continuous variables, smoothers with different degrees of freedom, as well as polynomials of order 2, were considered. Polynomials place parametric constraints on the shape, and because of the problems with collinearity in using higher order polynomials, only polynomials of order 2 were considered. Under the GAM environment, polynomials are fitted as if under the GLM framework; consequently, issues such as residual degrees of freedom, standard-error bands, and tests of significance are straightforward under a Poisson family.

Once the GAMs or GLMs were fitted, informal verification of the goodness of fit was obtained through plots of residuals and estimated standard errors. Figure 3 is an example of a diagnostic plot supplied in S-PLUS and used extensively in this analysis. The solid line on the figure is the estimated smoother $s(iso17, 3)$ fitted to the black-footed albatross take data where $iso17$ is the scaled distance to the 17°C surface isotherm multiplied by -1 if the observation was south of the isotherm; only observations with latitudes greater than 18.25°N were included. The dashed lines lie approximately two standard errors away from the central curve on either side and give a rough indication of the level of variability around the fitted curve. Even under Poisson assumptions and independence, the calculation of the standard errors in S-PLUS involves some crude approximations with unknown properties (Chambers and Hastie, 1993); hence, it is recommended that these bands be used only for diagnostic purposes. Also, these bands should not be interpreted as confidence bands as there has been no adjustment for bias and the normality assumption is questionable. In Figure 3, the standard error band displays a clear curvature that follows the curvature of the fitted curve but flares out near the endpoints of the range of observations. The fact that we cannot draw a horizontal line across the plot without going outside the band provides evidence that the distance to the 17°C isotherm was associated with black-footed albatross take.

The black circles in Figure 3 represent partial deviance residuals; these are simply the fitted term plus the deviance residual (Chambers and Hastie, 1993). The deviance is a function of the differences in the log-likelihoods and provides a measure of discrepancy between the fitted model and the saturated model; i.e., between the model and the data. The deviance residual for observation y_i is the square root of its contribution to the overall deviance multiplied by $sign(y_i - \hat{\mu}_i)$; i.e., it is positive if y_i is greater than its fitted value $\hat{\mu}_i$ and negative if y_i is less than $\hat{\mu}_i$. For our data, zeros dominated the responses and

$\hat{\mu}_i$ was consistently near zero; therefore, responses with 0 takes have negative deviance residuals, and responses of 1 or greater have positive deviance residuals and greater absolute values. A satisfactory diagnostic plot typically has residuals distributed evenly and randomly above and below the fitted curve. Again, due to the nature of our data, the curve will be closest to the zero values, and symmetry is not expected since zero is a lower boundary for counts. A healthy fit is indicated by the positive residuals following the pattern of the curve; i.e., the distance between the curve and the positive residuals is not distinctively greater in one section of the curve than in another.

Parts of the plot where the standard error band is particularly wide suggest that there are problems with the fit. Poor fits are most likely to be caused by sparse data or by a choice of degrees of freedom of the smoother that is too low to give an adequate representation of the relationship. Sparse data can be detected using the rug-plot along the bottom of Figure 3. The rug-plot gives the frequency of the x-values (Chambers and Hastie, 1993); where the frequency is high the rug-plot is a solid block. The band widens at the end of the range of distances because the reliability of a GAM fit is always reduced at the endpoints of the range (Hastie and Tibshirani, 1990). If the diagnostic plot indicates that the data are too sparse for the fit to be satisfactory, an alternative model formulation based on other predictors might be considered. The standard error bands in Figure 3 are wider on the left-side of the range, this may reflect problems with estimation when the event being modeled is very rare.

Interactions with nonparametric smooth terms are not fully supported in S-PLUS. In S-PLUS an interaction term can be modeled two dimensionally, $s(x_2, x_2)$, using the locally weighted regression smoother, but the numerical behavior of these estimates is not clear. In our case there appeared little advantage to using GAMs, and once this was determined, GLMs were used to fit the models and test for interactions. Likely two-way interactions were introduced into the models and the stepwise procedure based on the GIC criterion was used in the same manner as for GAMs. None of the interactions tested appeared to be associated with albatross take. We also compared models assuming overdispersed Poisson and negative binomial variates using disgnostic plots. The fits of the negative binomial models were tighter; therefore, we used the negative binomial model to predict total take.

To better understand how sets within the same trip may be correlated, we explored whether the takes were autocorrelated. To investigate autocorrelation, we created 12 new variables for each observation. The first six new variables were the observed black-footed and Laysan albatross takes and their sum for the previous two sets. For the first order regressive term (take on the previous set), the new variable was fixed at zero for the first set of the trip, and for the second order regressive term (the take two sets back), the new variables were fixed at zero for the first and second sets of the trip. The other six new variables were similar to the first six but were transformed to a binomial variable: a positive albatross take in a set was transformed to the value 1 indicating a 'take' versus 0, 'no take.' The first order regressive terms appeared to be significant even when added to the selected model; however, for both species, the binomial transformation of the first order regressive term of the species of interest appeared to explain the variation of take better than the actual number of previous takes. However, due to lack of time, we spent

very little time with model diagnostics to confirm which autoregressive terms were the best predictors. What we did learn, and what is important to note, is that there is evidence that sets within a trip are not independent and the correlation between sets within a trip is probably not constant but somewhat dependent on what has happened in the past sets.

4. Prediction

Once the predictive model was selected, the take for all unobserved sets was predicted by substituting in the values recorded in the logbook for all predictors in the model, and the expected take was used as the point estimate,

$$\hat{Y}_i = \hat{\mu}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}),$$

where x_{ij} was the logbook value for regressor j on set i. The point estimate for total predicted take $\hat{Y}_+$ was the sum of predicted takes for unobserved sets added to the sum of takes for observed sets,

$$\hat{Y}_+ = \sum_{i=1}^{n} Y_i + \sum_{i=1}^{N-n} \hat{Y}_i,$$

where n was the number of observed sets, and N was the number of longline sets recorded in the logbooks.

4.1. **Approximating Prediction Intervals and Correcting for Bias.** To measure the uncertainty in the point estimates, prediction intervals were approximated. These are similar to confidence intervals but wider. Confidence intervals measure the uncertainty in the parameter estimates and are used when interval estimates for the mean response $\beta\mathbf{x}$ are required. When model-based predictors are used to predict a total, we assume that the total is a random variable. Prediction intervals take into account both the inaccuracy of parameter estimates and the random fluctuations in the unobserved takes around the mean, $\mu_i = \beta\mathbf{x}$. For example, suppose the average take of albatross per set is estimated as $\hat{\mu} = 1$. We do not really expect $\hat{\mu}$ to equal the true average, so we calculate an interval, known as a confidence interval, that is expected to contain the true value of μ. The probability that a confidence interval will enclose μ is called the confidence coefficient. Now suppose that the true value of $\mu = 1$. We do not expect the take to equal 1 for every set but expect some takes to be 0, others 1, and some greater than 1: this is what is meant by the variation around the mean.

Instead of assuming an error structure, we used the bootstrap method to mimic the error structure of the original data. Prediction intervals were approximated using a bootstrapping algorithm based on that proposed by Davison and Hinkley (1997) but including an adaptation of block resampling. This resampling algorithm produces response variation in addition to variation in parameter estimates. The basic algorithm was as follows. For r=1,...,R, we created a bootstrap response y_i^* at $\mathbf{x_i}$ by

$$y_i^* = \hat{\mu}_i + \epsilon_i^* \sqrt{\hat{\mu}_i + \frac{\hat{\mu}_i^2}{\hat{\nu}}} \qquad (i = 1, ..., n),$$

where the ϵ_i^*'s were generated using Pearson residuals,

$$\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \frac{\hat{\mu}_i^2}{\hat{\nu}}}}.$$

As suggested by Davison and Hinkley (1997), stratification was necessary to create homogenous groups. We had better success in mimicking the original error structure when stratifying by the fitted values versus the residuals. It is clear from the residual plot (Fig. 4) that the smaller fitted values had relatively larger residuals than the larger fitted values. This is because an original count of 1 is large when compared to a fitted value of 0.001. If the bootstrap procedure assigned a large residual to a large fitted value, the result would be an extremely large, unrealistic generated value of take; similarly, if only small residuals were applied to small fitted values, we would not recreate the occasional positive take in the midst of zeros.

Even after stratifying the residuals, there was evidence that we had not accomplished mimicked the error structure. This is not surprising considering the evidence concerning the presence of autocorrelation between sets within a trip. To introduce this correlation, we used an adaptation to block resampling (Davison and Hinkley, 1997). The block resampling is often used when bootstrapping time-series data. The idea of the block resampling is to capture the autocorrelation structure by sampling, with replacement, blocks of consecutive observations. Since takes from sets within a trip were likely correlated, trip was defined as the sampling block. Within each stratum, the new ϵ_i^*'s were generated by randomly selecting a trip, with replacement, and then pasting selected trips from end to end to form a new series. Trips would be selected until the required number of ϵ_i^*s were generated and the new y_i^*s calculated. We then refitted the model with the y_i^*s and computed predicted totals $\hat{\mu}_{yr}^*$ for $yr = 1994, \cdots, 1999$. Finally, for each year we calculated the sum of generated observed y_{yr}^* and calculated the statistic

$$d_{yr}^* = \frac{y_{yr}^* - \hat{\mu}_{yr}^*}{\sqrt{\hat{\mu}_{yr}^* + \frac{\hat{\mu}_{yr}^{*2}}{\hat{\nu}}}},$$

where $\hat{\nu}$ was the estimated value from the original fit.

To approximate the prediction intervals, the R values of d_{yr}^* were ordered to give $d_{yr,(1)}^* \leq \cdots \leq d_{yr,(R)}^*$. The prediction limits were calculated as

$$(\hat{\mu}_{yr} + d_{yr,((R+1)*.025)}^* \sqrt{\hat{\mu}_{yr} + \frac{\hat{\mu}_{yr}^2}{\hat{\nu}}}, \hat{\mu}_{yr} + d_{yr,((R+1)*.975)}^* \sqrt{\hat{\mu}_{yr} + \frac{\hat{\mu}_{yr}^2}{\hat{\nu}}}).$$

For approximating intervals, it is recommended that $R \geq 999$; $R = 999$ was used in all bootstrap approximations reported here.

One drawback of this algorithm was that y^* could take on negative and noninteger values. To fix this, a constant was added to y^* and the new value was rounded to the nearest non-negative integer. The constant was selected so that the average total take of the bootstrap-generated data sets was comparable to the observed total take. The appropriateness of this algorithm was confirmed using the diagnostics suggested in Davison and Hinkley (1997). An advantage of this algorithm is that it implicitly adjusts for bias in the bootstrap distribution; however, the point estimate is not adjusted for

bias. Because there were a few odd trips with high takes, we were concerned about the bias they may have introduced. To correct for bias, we estimated bias as the bootstrap mean of $\hat{Y}^*_{yr} - \hat{Y}_{yr}$ for $yr = 1994, \cdots, 1999$, where $\hat{Y}^*_{yr}$ is the bootstrap-predicted total and $\hat{Y}_{yr}$ is the original predicted total. To adjust the point estimates for bias, we subtracted the estimated bias for each year from the original point estimate.

To attain the prediction intervals, including correcting for bias, we relied on heuristics and have not confirmed that we have achieved the 95% nominal level. Although we did check our methodology using simple diagnostic tests, because of time restrictions, we have not exhausted checking for all potential problems. Further research is required before we can confidently accept a resampling methodology for approximating 95% prediction intervals.

5. Results

Some of the variables considered and found to be associated with albatross take were poorly represented in the logbooks from 1994 and 1999 and thus were not considered for prediction. For both species, the negative binomial model appeared to provide the better fit. The two variables "distance to the approximate 17°C surface isotherm" and "a proxy of anticipated albatross density at location" were used in both models but in slightly different form. Table 2 summarizes how each predictor variable was expressed in the model. The maximal and minimal distance to the 17°C isotherm and anticipated density recorded in the observed sets were close to the extreme values in the logbooks, and between the two extremes the range of values were well represented in the observed sets. When modeling black-footed albatross takes, only sets with latitudes north of 16°C were fitted, and when modeling Laysan albatross takes, only sets above 15°N were fitted. Sets below these latitudes were assigned the minimal fitted value predicted by the model. Below these latitudes all observations had zero take; 18.25°N and 16.8°N were the southernmost locations at which a positive black-footed and Laysan albatross take was recorded in the observed sets, respectively. Figures 5 and 6 are plots of observed sets relative to the scaled distance to the 17°C isotherm and scaled anticipated density; sets with positive takes are elevated above the sets with no observed takes. For a given distance to the isotherm, there are more sets with observed positive takes, and the number of takes are generally higher closer to the nesting colonies.

Plots of annual take for each species are given in Figures 7 and 8. The shape of each plot represents the shape of the smoothed 'de-studentized' bootstrap distribution for estimated take. The shape is mirrored around a line representing the prediction interval with the point estimate represented as a square. We also computed point estimates of predicted take for the comparable Poisson GLM (log-linear model), a Poisson GAM with a smoother of 3 d.f. for the 17°C isotherm and linear term for anticipated density, and if we oversmoothed, a Poisson GAM with a smoother of 6 d.f. for the 17°C isotherm and a smoother of 4 d.f. for anticipated density. These results are tabulated in Tables 3 and 4. Although the GAMs allowed more flexible curves, the estimates are very similar to those derived from the GLMs. The largest difference in the estimates is the result of the fit in the right boundary where we have less information. The negative binomial appeared to fit the best; hence, our estimates are based on this model. Comparing

among years, the estimated take for 1994 is noticeably larger than for the other years but is then followed by a dramatic drop with 1995 having the smallest estimated total. After 1995, the estimated totals appear to stablilize. Out of the 765 sets with the highest predicted take (about 1.07% of all sets), 55% are from 1994; the remainder of the 765 sets are distributed fairly evenly (7% to 11%) among the other years. In 1998 and 1999 there were fewer sets in the highest 25% of predicted take values than in 1995, but because there were more sets in 1998 and 1999, the estimated total for 1995 is smaller. Compared to 1996 and 1997, 1995 had fewer sets in the highest 25% of the predicted values. Comparing among estimators, the estimates for 1994 have the most variation. This is because a large number of unobserved sets are at the right-end of the $iso17$ range where the rug-plot shows few observations and the standard error bands are wider.

6. Discussion

Several different avenues could be taken to improve the estimates of albatross take. First, there are trips where the observed takes were very high compared to other trips. Robust model fiitting and estimation would ensure that the influence of these trips was not disproportionate or misleading. This is an area that needs further research. Second, the methodology for approximating prediction intervals needs to be strengthened. Third, modeling the dependence structure among observations using an autoregressive, generalized linear mixed model (GLMM), or a combination of these two, has the potential of increasing precision. There are several different ways to construct a GLMM, but without further research, it is hard to know which of these methods are preferable or even appropriate for modeling albatross take. Also, how to predict takes using an autoregressive model for the unobserved trips is not clear. Fourth, if not using a mixed model, different scatterplot smoothers and methodologies for fitting GAMs should be compared. Because we are modeling a rare event with dependent errors and related explanatory variables, the best methodology for selecting and fitting a GAM is very unclear. Because of the software and time required we have not spent much time on this issue. Unfortunately, the methodology for doing mixed models in the GAM environment has yet to be developed.

The estimated predicted albatross takes for 1994-98 changed from years past (Table 5) when regression trees were used to predict take (Laurs et al., 1999). The methodology for modeling take was not changed arbitrarily. The theory used to construct regression trees is similar to that of linear regression, and the take data do not meet the underlying assumptions. Previously, with the turtle data, the models selected by regression trees were compared to the models selected by classification trees and the overdispersed Poisson model (McCracken, 2000). If in fact regression trees were robust to the departures from the underlying assumption present in the take data, we would expect similar results concerning model selection and prediction, but this was not the case.

It cannot be stressed strongly enough that since the observer program was an observational study, only associations between predictor variables and take can be identified. Causality cannot be concluded without further evidence. Furthermore, the models can be used only to predict the events that were observed within the range of the data. To apply observational results to predict future takes requires additional assumptions about the

future behavior of predictors compared to the behavior of the existing data (Weisberg, 1985).

It is easy when using a simple regression model, $Y = \beta_0 + \beta_1 X + \epsilon$, to form the bad habit of referring to β as the average change in the response *due to* a unit increase in X. Many different reasons may contribute to a prediction variable having a relationship with the response variable, and the selection routines and regression calculations give few, if any, clues to the explanation (Snedecor and Cochran, 1980). Sometimes X itself is not causally related to Y but is related to a variable Z that is causally related to Y. For example, when matches were commonly used to light cigarettes, it would not have been surprising to discover a relationship between the average number of matches a person carried and the risk of developing lung cancer. With our current understanding concerning the relationship between smoking and cancer, concluding that carrying matches caused cancer and carrying fewer would reduce the risk of cancer would probably be considered ludicrous.

Interpretation of a multiple linear regression model , $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$, is more difficult. Before concluding that β_i $(i = 1, \ldots, p)$ is the average change in Y *due to* a unit increase in X_i, we also need to assume that a predictor can in fact be changed by one unit without affecting the other predictors and that the model fitted to the available data will apply when the predictor is so changed. This is hardly ever the situation with observational data. When the predictor variables are highly intercorrelated, the value and even the sign of the regression coefficients can change according to the other predictor variables that are in the model. Hence, our inability to estimate effects accurately from observational data is not resolved by simply measuring more and more potential predictors until eventually all important predictors have been measured (Robins and Greenland, 1986). If many predictor variables are strongly associated with the response Y, even to conclude causality from a case-control or matching observational study, we need to have nearly correct prior beliefs concerning what variables related to Y are truly important and also the magnitude of their effect (Robins and Greenland, 1986).

In terms of the albatross data, latitude and fish target species are highly correlated, but latitude is likely a proxy for target species and the spatial distribution of albatross density, food supply, and many other environmental variables that have not been observed. Latitude is the angular distance on a meridian north or south of the equator, and it is hard to see how this distance could actually be the cause of increased takes. However, it does seem reasonable that the numbers of albatross within a certain distance of the vessel and possibly the fishing practices being employed could be causal factors in take. Latitude is related to both of these variables. If latitude is included in the model, it is probably because it is related to a causal variable. To use this model to predict the reduction of take if fishing is closed between certain latitudes is likely to provide misleading results. Even if indeed latitude and target species do have causal relationships with take and are included in the model, we have no measure of the magnitude of their separate effects. If an area closure results in fishermen maintaining the same practices but just relocating, using this model to predict the reduction in takes is likely to provide misleading results. Until there is a better understanding of the causality of higher

take rates, concluding causality solely from the observer data is beyond the capacity of statistics.

REFERENCES

Atkinson, A. C.
1980. A note of the generalized information criterion for choice of a model. Biometrika 67, 413-418.

Atkinson, A. C.
1981. Likelihood ratios, posterior odds and information criteria. J. Econometrics 16, 15-20.

Chambers, J. M., and T. J. Hastie.
1993. Statistical Models in S. Chapman and Hall, New York.

Cox, D. R.
1983. Some remarks on over-dispersion. Biometrika 70, 269-74.

Davison, A. C. and D. V. Hinkley.
1997. Bootstrap Methods and their Application. Cambridge University Press, New York.

Firth, D.
1987. On the efficiency of quasi-likelihood estimation. Biometrika 74, 233-45.

Hastie, T. J. and R. J. Tibshirani.
1990. Generalized Additive Models. Chapman and Hall, New York.

Hurvich, C. M., J. S. Simonoff, and C. L. Tsai.
1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, J. R. Statistic. Soc. B, 60:271-293.

Laurs, R. M. and Associates.
1999. 1999 Program Review Honolulu Laboratory. Honolulu Lab., Southwest Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, Honolulu, HI 96822-2396, Southwest Fish. Sci. Cent. Admin. Rep. H-99-05, 110 p.

McCracken, M. L.
2000. Estimation of sea turtle take and mortality in the Hawaiian longline fisheries. Honolulu Lab., Southwest Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, Honolulu, HI 96822-2396. Southwest Fish. Sci. Cent. Admin. Rep. H-00-06. 29 p.

McCullagh, P.
1983. Quasi-likelihood functions. Ann. Statist. 11: 59-67.

McCullagh, P. and J. A. Nelder.
1989. Generalized Linear Models, 2nd edition. Chapman and Hall, New York.

Robins, J. M. and S. Greenland.
1986. The role of model selection in causal inference from nonexperimental data. Am. J. of Epidemiol. 123: 392-402.

Schimek, M. G. and B. A. Turlach.
2000. Additive and generalized additive models.
In: Smoothing and Regression Approaches, Computation, and Application, ed. Schimek, M. G., Wiley, New York, pp. 277-327.

Snedecor, G. W. and W. G. Cochran.
1980. Statistical Methods, 7th edition. Iowa State University Press, Ames.

Statistical Sciences Inc.
1993. S-PLUS reference manual. StatSci, a division of Mathsoft, Inc., Seattle.

Thurston, S. W., M. P. Wand, and J. K. Wiencke.
2000. Negative binomial additive models. Biometrics, 56, 139-144.

Venables, W. N. and B. D. Ripley.
1999. Modern Applied Statistics with S-PLUS, 3rd Ed., Springer, New York.

Weisberg, S.
1985. Applied Linear Regression, 2nd Ed., Wiley, New York.

TABLE 1. Explanatory variables considered for predicting total albatross take. All continuous variables were scaled according to Equation 3.2.

	Variable	Notes
Location in time and space	latitude (lat)	degrees north
	longitude (long)	degrees east
	distance to $17°C$ isotherm	calculated from lat, lon, and sst
	distance to $19°C$ isotherm	calculated from lat, lon, and sst
	distance to nesting colonies	colonies considered: Kure Atoll, Midway Atoll, Laysan Island, Pearl and Hermes Reef, Necker Island, Lisianski Island, French Frigate Shoals, Nihoa Island, Kauai, Niihau, Kaula
	proxy for anticipated albatross density at location	see Equation 3.1
	year	1994-99, looked at pooling categories
	month	January-December, looked at pooling categories
	day	represented as a circular variable for a year using the cosine and sine functions
Condition of gear	hooks	
	hooks/float	
Environment	temperature	sea surface temperature (sst)
Catch of other species		total and proportion of total
	yellowfin	
	skipjack	
	albacore	
	swordfish	
	blue shark	
	mahimahi	
	striped marlin	
	blue marlin	
	wahoo	
	spearfish	
	opah	
Other	vessel length	registered length
	trip type	3 categories (swordfish,tuna,mixed)

TABLE 2. Explanatory variables included in the prediction models

Species	Explanatory variables
Black-footed	distance to 17°C isotherm as a polynomial $iso17 + iso17^2$ proxy for anticipated albatross density at location
Laysan	distance to 17°C isotherm proxy for anticipated albatross density at location

TABLE 3. (a) Black-footed albatross take estimates with 95% prediction intervals for 1994-1999; (b) point estimate before adjusting for bias; (c) point estimates predicted from the corresponding overdispersed Poisson model (not adjusted for bias); (d) point estimates predicted from a Poisson GAM that includes a smoother with 3 d.f. for the 17°C isotherm and a linear function for anticipated density (not adjusted for bias); and (e) point estimates predicted from a Poisson GAM that includes a smoother of 6 d.f. for the 17°C isotherm and a smoother of 4 d.f. for the anticipated density (not adjusted for bias).

Black-footed albatross takes					
Year	a	b	c	d	e
1994	1830 [1457-2239]	2024	1903	1918	1895
1995	1134 [899-1376]	1192	1190	1209	1239
1996	1472 [1199-1811]	1415	1389	1409	1407
1997	1305 [1077-1592]	1311	1290	1310	1301
1998	1283 [1028-1601]	1338	1311	1334	1327
1999	1301 [1021-1600]	1281	1258	1260	1269

TABLE 4. (a) Laysan albatross take estimates with 95% prediction intervals for 1994-1999; (b) point estimate before adjusting for bias; (c) point estimates predicted from the corresponding overdispersed Poisson model (not adjusted for bias); (d) point estimates predicted from a Poisson GAM that includes a smoother with 3 d.f. for the 17°C isotherm and a linear function for anticipated density (not adjusted for bias); and (e) point estimates predicted from a Poisson GAM that includes a smoother of 6 d.f. for the 17°C isotherm and a smoother of 4 d.f. for the anticipated density (not adjusted for bias).

Laysan albatross takes						
Year	a		b	c	d	e
1994	2067	[1422-2948]	2148	1743	1663	1629
1995	844	[617-1131]	859	837	868	895
1996	1154	[835-1600]	1081	985	991	1001
1997	985	[715-1364]	965	899	885	868
1998	981	[679-1360]	975	933	958	1021
1999	1019	[688-1435]	1009	936	995	1073

TABLE 5. Black-footed and Laysan albatross regression tree take estimates with 95% confidence intervals for 1994-1998 (Laurs et al., 1999).

	Black-footed albatross		Laysan albatross	
Year	Estimated total take	95 %CI	Estimated total take	95 %CI
1994	1994	[1508-2578]	1828	[933-2984]
1995	1979	[1439-2497]	1457	[767-2308]
1996	1568	[1158-1976]	1047	[569-1610]
1997	1653	[1243-2102]	1150	[599-1875]
1998	1653	[1479-2470]	1479	[822-2336]

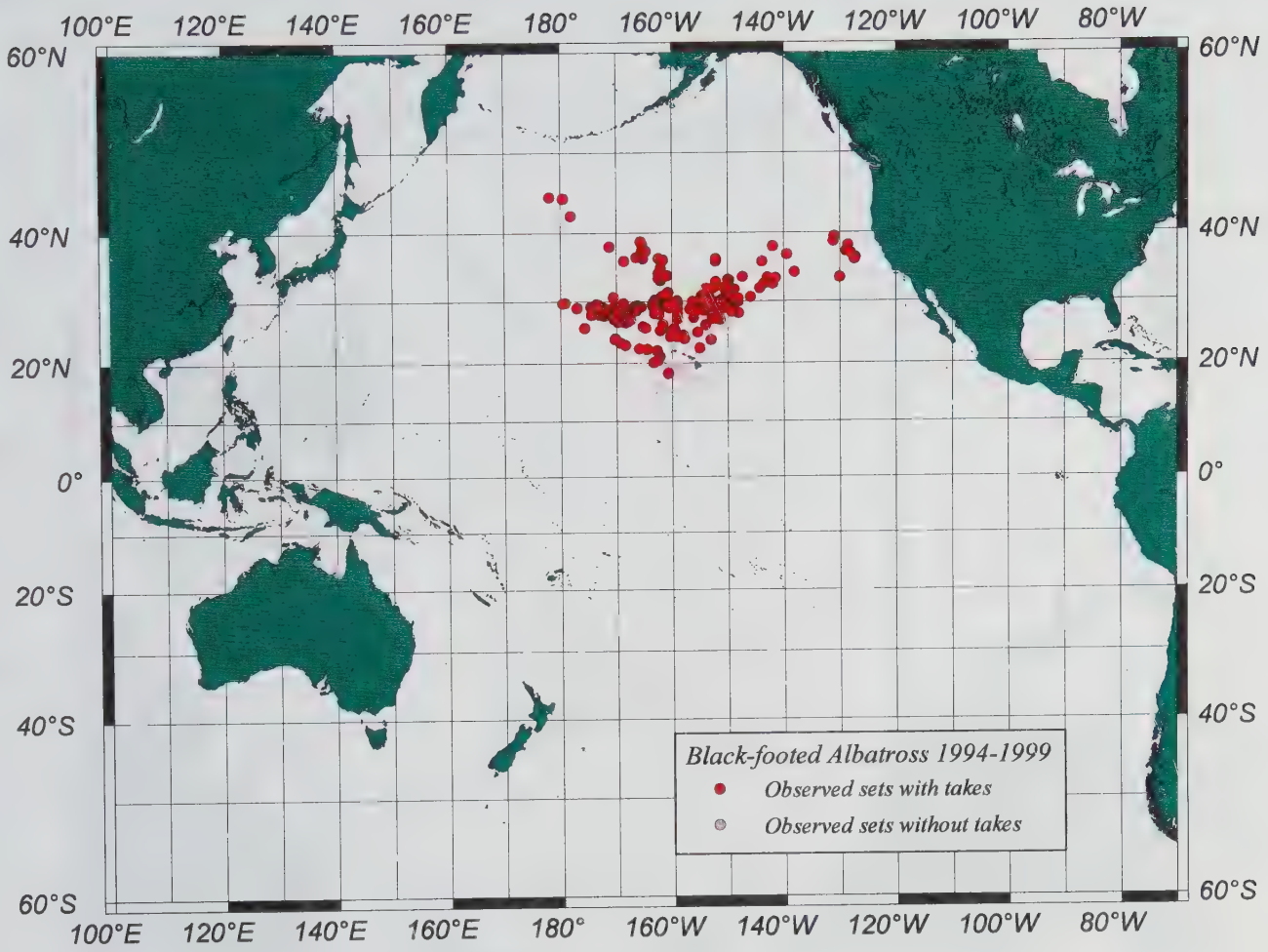


Figure 1. Approximate geographical positions of observed sets with positive black-footed albatross takes represented as red circles and takes of zero represented as pink dots. The Hawaiian Islands and other land masses are also shown.

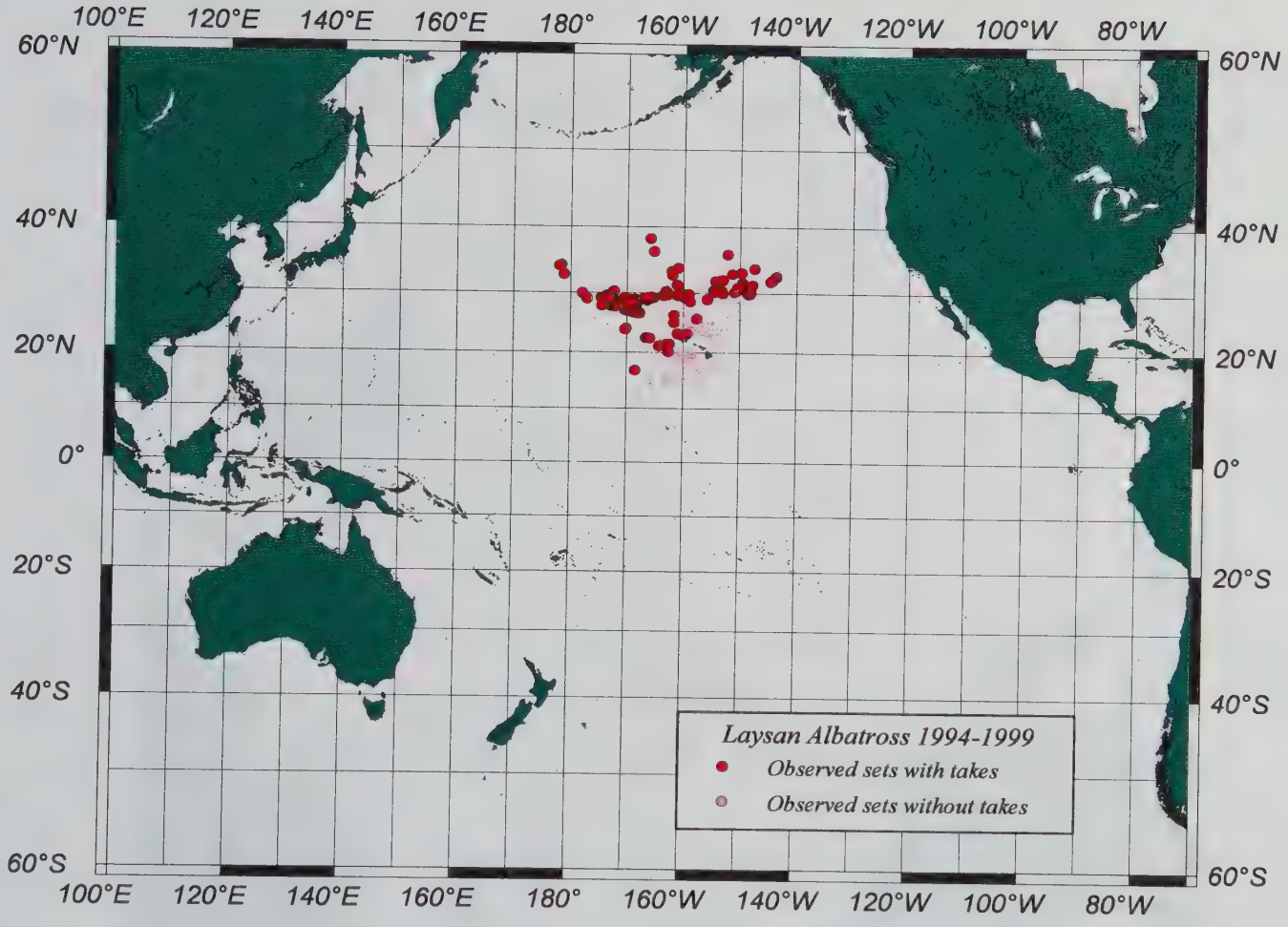


Figure 2. Approximate geographical positions of observed sets with positive Laysan albatross takes represented as red circles and takes of zero represented as pink dots. The Hawaiian Islands and other land masses are also shown.

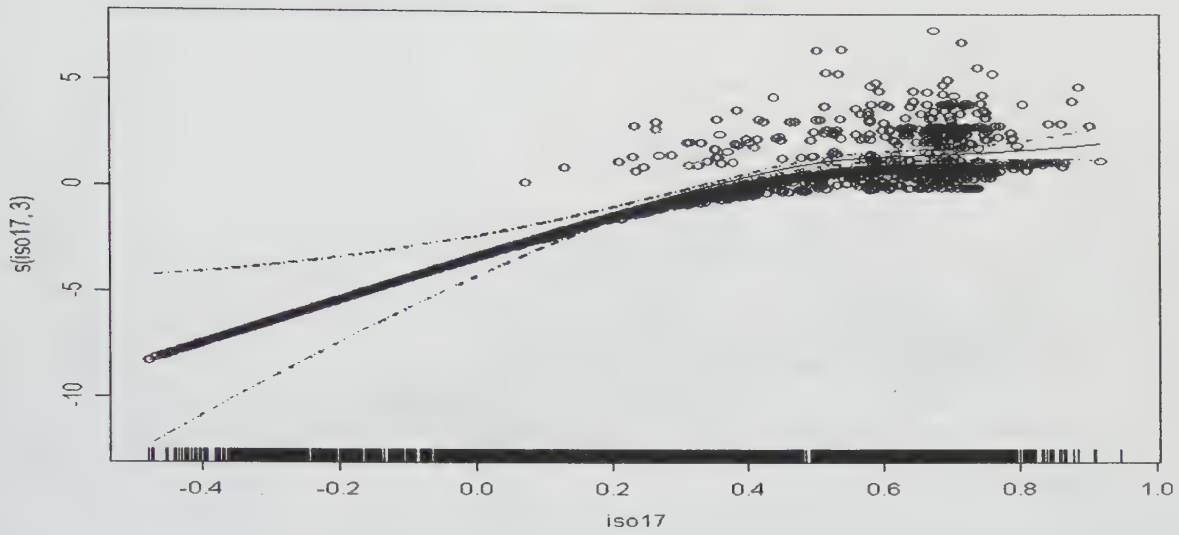


Figure 3. The fit of black-footed albatross take to the scaled distance to the 17°C isotherm, for latitudes greater than or equal to 16°N. The solid line represents the fitted smooth curve with 3 degrees of freedom, the dashed lines denote the fitted smooth plus or minus 2 standard errors (approximate) and demarcate a "standard error band", the black circles represent partial deviance residuals, and the bars on the x-axis are the rug-plot. The residuals are well distributed above and below the curve and follow the basic line of the curve. The standard error band shows a definitive curve and is narrow in the center of the curve, but wider at the endpoints where there are few observations.

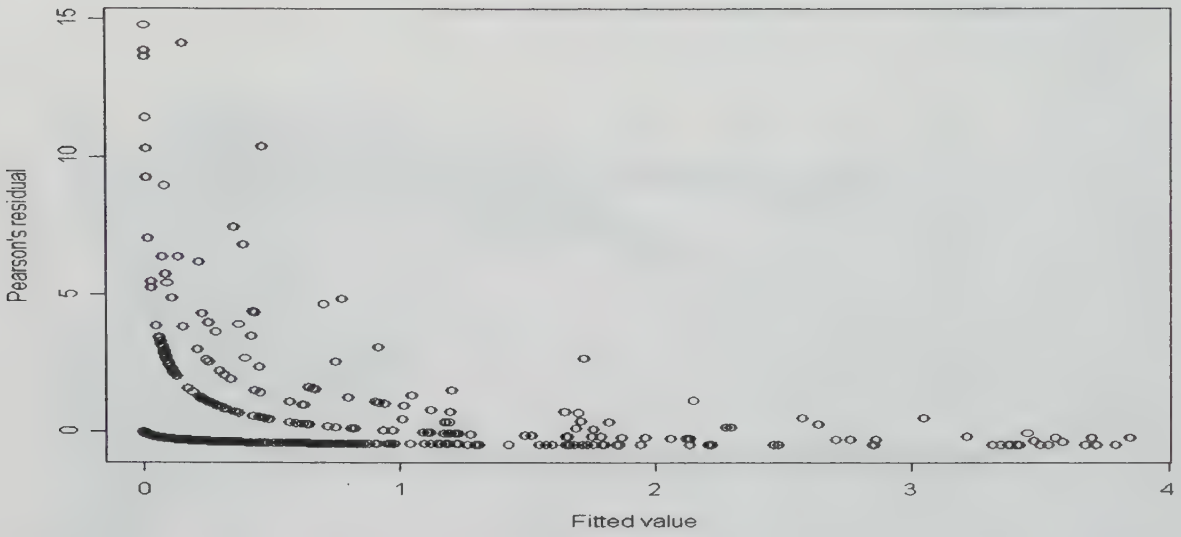


Figure 4. The fitted values of Laysan albatross takes are plotted against their corresponding Pearson's residual.

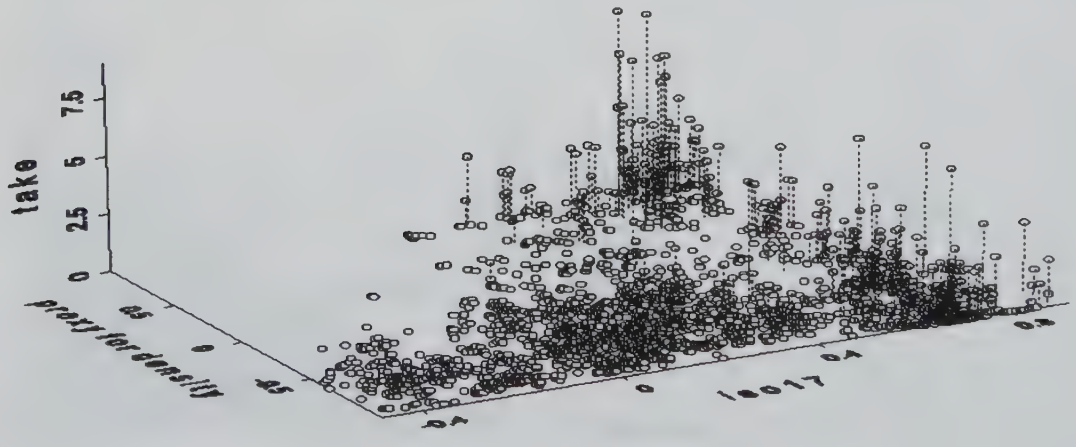


Figure 5. Takes of black-footed albatross versus the distance to the 17°C isotherm (iso17) and the proxy for anticipated black-footed albatross density at location (truncated data).

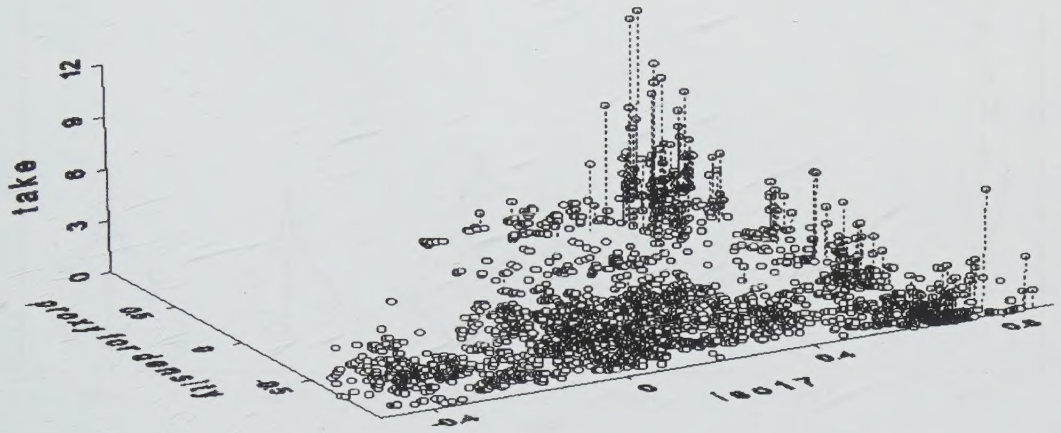


Figure 6. Takes of Laysan albatross versus the distance to the 17°C isotherm (iso17) and the proxy for anticipated Laysan albatross density at location (truncated data).

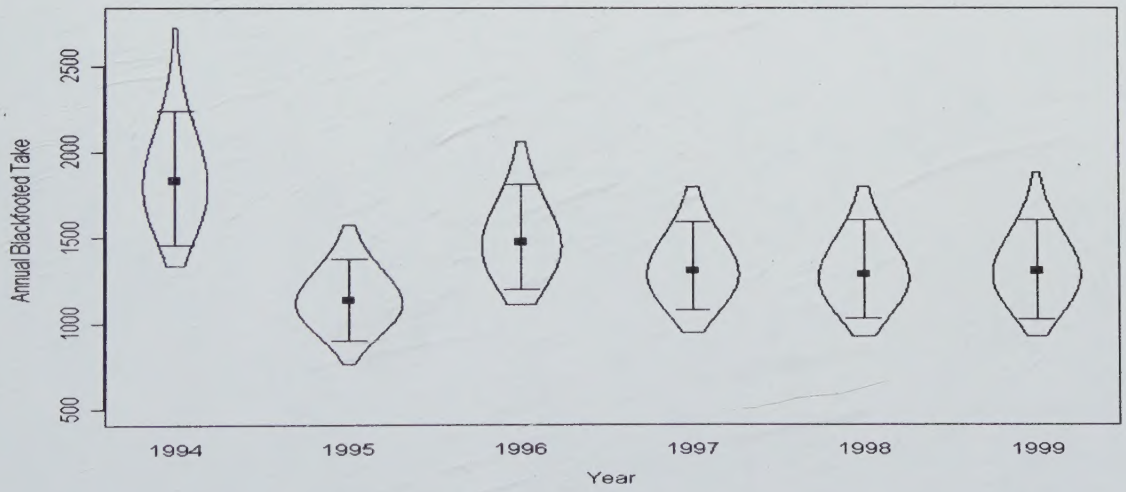


Figure 7. Take estimates for black-footed albatross. The approximate distributions of estimated take are mirrored around the prediction intervals (dark line with bars at each end). The dark squares in the plots represent the point estimates.

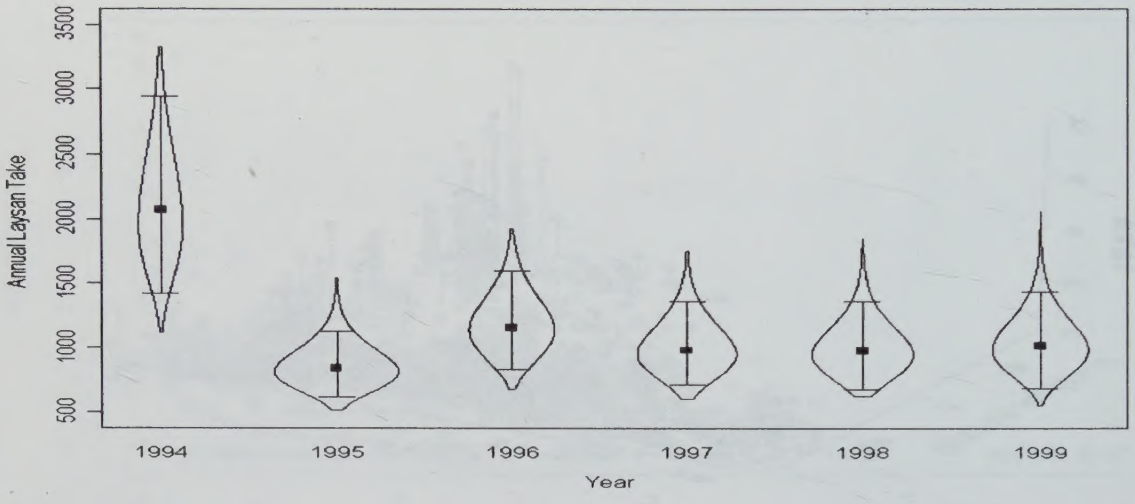


Figure 8. Take estimates for Laysan albatross. The approximate distributions of estimated take are mirrored around the prediction intervals (dark line with bars at each end). The dark squares in the plots represent the point estimates.